Modelling Limit Order Book Dynamics Using Poisson and Hawkes Processes

Yuanda Chen
Major Advisor: Dr. Alec N. Kercheval

Department of Mathematics
Florida State University

May 22, 2013
Outline

1. Introduction
2. Poisson processes
3. Modelling order book dynamics
4. Hawkes processes
5. Future Research
Introduction

1. From quote-driven to order-driven markets.
3. A limit order sits in the order book until it is either executed against a market order or canceled.
4. Three things can happen:
   (a) new order added
   (b) existing order executed against market orders
   (c) existing order canceled
5. High frequency trading and data make modelling at transaction level possible.
Counting processes

**Definition (Point processes)**

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{t_i\}_{i=0,1,2,...}$ be a sequence of non-negative random variables such that for all $i$, $t_i < t_{i+1}$ and $t_i \rightarrow \infty$ a.s.. We call $\{t_i\}$ a point process on $\mathbb{R}^+$. 

**Definition (Counting processes)**

The right-continuous process

$$N(t) = \sum_{i=0,1,2,...} 1_{t_i \leq t < \infty}$$

a.s. is called the counting process associated with $\{t_i\}_{i=0,1,2,...}$. 
A counting process satisfies:

1. $N(t) \geq 0$.
2. $N(t)$ is integer valued.
3. If $s < t$, then $N(s) \leq N(t)$.
4. For $s < t$, $N(t) - N(s)$ is the number of occurrence in $(s, t]$. 

Yuanda Chen
PhD Candidacy Examination
May 22, 2013
Poisson processes

Definition (Poisson processes)

The counting process \( \{N(t), t \geq 0\} \) is said to be a Poisson process with rate \( \lambda, \lambda > 0 \), if

(i) \( N(0) = 0 \).

(ii) The process has independent increments.

(iii) The number of events in any interval of length \( t \) is Poisson distributed with mean \( \lambda t \). That is for all \( s, t \geq 0 \)

\[
P\{N(t + s) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots
\]

Proposition

The interarrival times \( T_n, n = 1, 2, \ldots \) for a Poisson process, are independent identically distributed exponential random variables with mean \( 1/\lambda \).
**Definition**

The function $f$ is said to be $o(h)$ if $\lim_{h \to 0} \frac{f(h)}{h} = 0$.

**Proposition**

The counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda$, $\lambda > 0$, if and only if

(i) $N(0) = 0$.

(ii) The process has stationary and independent increments.

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(iv) $P\{N(h) \geq 2\} = o(h)$.

as $h \to 0$. 
Consider a price grid \( \{1, 2, \ldots, n\} \) of a range of prices at integer multiples of a price tick, out of which prices are unlikely to appear for a short period. At a given time \( t \), the state of the order book is defined as a vector associating with the number of outstanding limit orders at each price on the grid:

\[
X(t) = (X_1(t), X_2(t), \ldots, X_n(t))
\]

where \( |X_p(t)| \) is the number of outstanding limit orders at price corresponding to \( p \) on the price grid. If \( X_p(t) < 0 \) then it means that there are \( -X_p(t) \) buy orders at price \( p \); if \( X_p(t) > 0 \), then there are \( X_p(t) \) sell orders at price \( p \).
Some crucial variable can be derived from the state of the order book.

(i) The ask price, or the best sell price,
\[ p_A(t) = \inf \{ p = 1, \ldots, n, X_p(t) > 0 \} \land (n + 1). \]

(ii) The bid price, or the best buy price,
\[ p_B(t) = \sup \{ p = 1, \ldots, n, X_p(t) < 0 \} \lor 0. \]

(iii) The mid-price,
\[ p_M(t) = \frac{p_B(t) + p_A(t)}{2}. \]

(iv) The bid-ask spread,
\[ p_S(t) = p_A(t) - p_B(t). \]

(v) The number of buy orders at a distance \( i \) from the ask,
\[
Q_i^B(t) = \begin{cases} 
X_{p_A(t)-i}(t) & 0 < i < p_A(t) \\
0 & p_A(t) \leq i < n 
\end{cases}
\]

(vi) The number of sell orders at a distance \( i \) from the bid,
\[
Q_i^A(t) = \begin{cases} 
X_{p_B(t)+i}(t) & 0 < i < n - p_B(t) + 1 \\
0 & n - p_B(t) \leq i < n 
\end{cases}
\]
Assumptions

All three events that can happen to the order book are assumed to be mutually independent, with arrival times modelled as Poisson processes with different rates:

1. Incoming limit orders: $\lambda(i)$ which has the form $\lambda(i) = k/i^\alpha$.
2. Incoming market orders: a constant $\mu$.
3. Cancellation of limit orders: $\theta(i)x$, proportional to the number of outstanding orders $x$. 

Yuanda Chen
PhD Candidacy Examination
May 22, 2013
Slide 10/24
If we define $x^{p\pm 1} = x \pm (0, \ldots, 1, \ldots, 0)$ where the 1 is at the $p$th component, then the transition of the state can be expressed as:

- $x \rightarrow x^{p-1}$: incoming limit buy order with rate $\lambda(p_A(t) - p)$ for $p < p_A(t)$.
- $x \rightarrow x^{p+1}$: incoming limit sell order with rate $\lambda(p - p_B(t))$ for $p > p_B(t)$.
- $x \rightarrow x^{p_A(t)-1}$: execution of best sell order with rate $\mu$.
- $x \rightarrow x^{p_B(t)+1}$: execution of best buy order with rate $\mu$.
- $x \rightarrow x^{p-1}$: cancellation of sell order with rate $\theta(p_A(t) - p)|x_p|$ for $p < p_A(t)$.
- $x \rightarrow x^{p+1}$: cancellation of buy order with rate $\theta(p - p_B(t))|x_p|$ for $p > p_B(t)$. 
Definition (Intensity function)

Let $N$ be a counting process adapted to a filtration $\mathcal{F}_t$. The left-continuous conditional intensity function is defined as (Daley 2003)

$$\lambda(t|\mathcal{F}_t) = \lim_{h \to 0} \frac{1}{h} P\{N(h + t) - N(t) > 0|\mathcal{F}_t\}.$$

Poisson processes revisited

$$\lim_{h \to 0} \frac{1}{h} P\{N(t + h) - N(t) > 0\} = \lim_{h \to 0} \frac{\lambda(t)h + o(h)}{h} = \lambda(t).$$
A univariate self-excited process $N(t)$ is given by allowing the intensity itself be stochastic, characterized using the following Stieltjes integral (Hawkes 1971)

$$\lambda(t) = \nu + \int_{-\infty}^{t} \gamma(t-u) dN_u = \nu + \sum_{t_i < t} \gamma(t-t_i),$$  \hspace{1cm} (1)$$

where $\nu > 0$ is called the *base intensity* and $\gamma(t) > 0$ the *excitation kernel* such that

$$0 < m = \int_{0}^{\infty} \gamma(u) du < 1.$$  \hspace{1cm} (2)$$
A simple version of Hawkes process $\lambda(t)$ is defined by using a constant base intensity and an exponential kernel $\gamma(t) = \alpha e^{-\beta t}$:

(i) $N(0) = 0$.

(ii) $\lambda(t) = \nu + \alpha \int_0^t e^{-\beta(t-s)} dN_s = \nu + \alpha \sum_{t_i < t} e^{-\beta(t-t_i)}$.

(iii) $P\{N(t + h) - N(t) = 1 | \mathcal{F}_t\} = \lambda(t) h + o(h)$.

(iv) $P\{N(t + h) - N(t) \geq 2 | \mathcal{F}_t\} = o(h)$. 
Lemma (Existence)

If $\nu > 0$ and $m < 1$ there exists a stationary Poisson cluster process with rate $\lambda = \nu/(1 - m)$ satisfying (1) and (2).

Lemma (Uniqueness)

There exists at most one stationary orderly point process of finite rate whose complete intensity is given by (1) and (2).
Definition (Poisson cluster processes)

A Poisson cluster process $X \subset \mathbb{R}$ is a point process such that:

(a) The immigrants (cluster centers) are distributed according to a homogeneous Poisson process $I$ with points $X_i \in \mathbb{R}$ and intensity $\nu > 0$.

(b) Each immigrant $X_i$ generates a cluster $C_i$ which is a finite point process (i.e., it has a finite number of points almost surely) containing $X_i$.

(c) Given the immigrants, the centered clusters

$$C_i - X_i = \{ Y - X_i : Y \in C_i \}, X_i \in I$$

are iid and independent of $I$.

(d) $X$ consists of the union of all clusters.
Existence and Uniqueness

Theorem

Hawkes (1974) If \( \nu > 0 \) and the non-negative function \( \gamma(u) \) satisfies

\[
0 < m = \int_{0}^{\infty} \gamma(u) du < 1
\]

then there is precisely one stationary orderly process of finite rate whose complete intensity function satisfies (1).
Simulation

Sample Path

Figure: An Example of the Intensity of a Hawkes Process
\((\nu = 1.2, \alpha = 0.6, \beta = 0.8)\)
The original thinning algorithm was developed by Lewis (1978) for simulating inhomogeneous Poisson processes with rate $\lambda(t)$.

1. Choose a constant $\lambda^*$ such that $\lambda^* \geq \lambda(t)$ for all $t$.
2. Generate a homogeneous Poisson process with rate $\lambda^*$.
3. Accept with probability $\lambda(\cdot)/\lambda^*(\cdot)$.

Ogata (1981) modified this algorithm to generate Hawkes processes.
Algorithm

1. Set \( n = 0 \) and \( s = 0 \) and set \( t_0 = s = 0 \).

2. Repeat until \( s > T \).
   
   (i) Compute \( m(s) = \lambda(s|t_1, \ldots, t_n) + \alpha \) if \( s \) was accepted as some \( t_i \) and \( m(s) = \lambda(s|t_1, \ldots, t_n) \) if \( s \) was rejected. This is the rate at which we generate the next point as if we are generating a Poisson process.

   (ii) Generate an independent exponential random variable \( \Delta s \) with mean \( 1/m(s) \). Then \( s + \Delta s \) is the next point waiting to be accepted/rejected.

   (iii) Generate an independent uniform random variable \( D \) and accept \( s + \Delta s \) by letting \( n = n + 1 \) and \( t_n = s + \Delta s \) if \( D \leq \lambda(s + \Delta s|t_1, \ldots, t_n)/m(s) \); otherwise, reject \( s + \Delta s \).

   (vi) Set \( s = s + \Delta s \).

3. Output \( t_0, t_1, \ldots, t_n \) as the occurrence times of the Hawkes process.
Residual Analysis

1. Published in (Ogata 1988).
2. Test how well the Hawkes process with known parameters fits the data.
3. Define the compensator: $\Lambda(t) = \int_0^t \lambda(u)du$.
4. $\Lambda(t_i) - \Lambda(t_{i-1}), i = 2, \ldots, n$ are i.i.d. exponential random variables with mean 1 (Daley 2003).
5. In our case,

$$\Lambda(t_i) = \int_0^{t_i} \lambda(u)du$$

$$= \int_0^{t_i} (\nu + \sum_{\{j:t_j<u\}} ae^{-b(u-t_j)})du$$

$$= \nu t_i + \frac{a}{b} \sum_{\{j:t_j<u\}} (1 - e^{-b(t_i-t_j)})$$
Figure: Residual Analysis of Simulated Hawkes/Poisson Process on [0, 1000]
Future Research

1. Run simulations
2. Prove propositions
3. Analytical solution
4. ...
References


